

Automatic Detection and Classification of Beluga Whale Vocalizations

Ramón Miralles^{*1}, Guillermo Lara², Alicia Carrión³, Jose Antonio Esteban⁴

Instituto de Telecomunicación y Aplicaciones Multimedia (iTEAM), Universitat Politècnica de València, Camino de Vera S/N, 46022, Valencia, Spain

Parques Reunidos Valencia S. A. L'Oceanográfico, Ciudad de las Artes y las Ciencias, Camino de las Moreras S/N, 46013, Valencia, Spain

^{*}rmiralle@dcom.upv.es; ²guilamar@iteam.upv.es; ³alcarga4@upv.es; ⁴investigacion@oceanografic.org

Abstract

In this work, an algorithm has been proposed for real time detection and classification of beluga whale calls. The detection algorithm is based on an adaptive activity detector that exploits a priori knowledge of the longest/ shortest beluga whale sound unit. Optimum parameter values of the proposed detector are obtained by simulation to maximize the difference between Detection Probability and False Alarm Probability. A set of features that allow successful classification by means of a Naive Bayes classification algorithm has been put forward as well. Three classification categories related to observed beluga behaviours were selected. In a different perspective, the proposed features can be employed to obtain clues of how beluga sounds are produced and the degree of control that the specie has over its internal organs. As an example, the presence of nonlinearities such as subharmonics and frequency jumps have been detected and related to some extracted features. This technique can serve as a complement to more rigorous studies based on video information and ultrasonic sensors for whale monitoring.

Keywords

Statistical Signal Processing; Signal Detection and Classification; Nonlinear Acoustics; Marine Bioacoustic Signal Processing

Introduction

In 1971, Payne and McVay defined the structure of humpback whale songs as themes that are repeated in specific patterns; and the building blocks are termed units (the shortest continuous sound between two silences) [15]. A relatively recent study shows that the vocalization patterns might be a useful tool to measure "animal welfare" in belugas (*Delphinapterus leucas*). In his work, Castellote demonstrates that during stress periods, such as that produced by veterinary manipulation or air transportation to new facilities, the vocalization rate and patterns change [12]. Other studies suggest that automatic systems for

continuously monitoring communicative beluga vocalizations could become an important tool for the research on animal behaviour and animal care [11]. The implementation of these detection systems is not an easy task and has never been done in belugas. The rich vocabulary of beluga whales, the complexity of their songs and the presence of interferences from different sources make it difficult to design automatic detectors.

The algorithms typically employed are not prepared to work in real time. Most of the proposed detectors are based on the computation of some kind of time-frequency representation. After that, the vocalization patterns are searched by means of different signal processing techniques: spectrogram correlation detector (XBAT) [3, 5], cross correlation with a matched filter kernel [3, 9], neural networks [3], etc. Although these approaches give quite accurate results, they are developed as research laboratory tools rather than for designing continuous monitoring systems. Additionally, the aforementioned algorithms need high computation capacity processors.

This work proposes an alternative to the problem of detection and classification of beluga whale vocalizations based on a reduced set of extracted features. This technique allows real time processing of the audio registers and designing automatic systems.

The remainder of this paper is structured as follows. In the second section, we present an adaptive activity detector and the selected features for classification. In the third section, we validate the proposed system in a controlled environment. In the forth section, we give some examples using the presented features to automatically detect the presence of nonlinearities. Finally, we present our conclusions and future work.

The research was carried out in the Oceanografic

facilities (Ciudad de las Artes y las Ciencias de Valencia) as part of a collaborative research project between the Institute of Telecommunications and Multimedia Applications and the Oceanographic biologists.

The Proposed Approach

The proposed approach to the problem is schematically described in Fig. 1. The system is composed of two main blocks: an adaptive activity detector and a feature-based classifier. Before explaining in detail different parts of the proposed algorithm, some assumptions on signal statistics should be done. It is assumed that the signal acquired by the hydrophone can be modeled as the sum of two stochastic processes: a quasi-stationary process that models the ambient noise and a quasi-stationary process that models the whale vocalization. These assumptions have been made after studying a large number of records provided by the Oceanographic researchers. They are quite accurate if the analysis window is short enough.

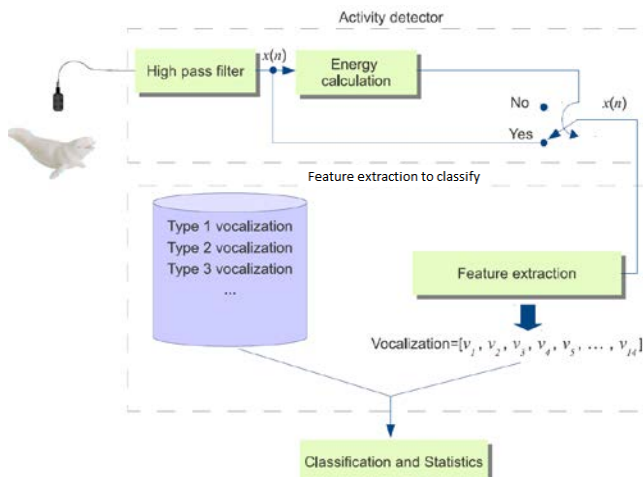


FIG. 1 PROPOSED AUTOMATIC DETECTOR FOR BELUGA WHALES VOCALIZATIONS

Adaptive Activity Detector

The first stage of the algorithm consists of a Finite Impulse Response digital high pass filter (1000 Hz cutoff frequency). This filter removes noise components out of the beluga frequency range and is proved to be an effective tool to remove low frequency interference sounds produced by other species that can be found at the Oceanographic facilities.

After this high pass filter, we employ an activity detector algorithm fine tuned for beluga whale calls. The proposed activity detector is similar to many long-established detectors applied in robust detection of

speech [8]. However, we have added some modifications to these detectors that take profit of the beluga whale songs structure.

Let us assume that the stochastic processes that model the ambient noise and beluga whale song picked up by the hydrophone are zero mean Gaussian processes of given variance. If noise and signal (beluga sounds) were zero mean white Gaussian stationary processes of different variances, the Neyman Pearson detector would be the energy detector [17]. Unfortunately, the noise component resembles more closely a quasi-stationary process than a stationary one. Noise variance changes slowly with time due to the ambient noise, for instance.

Using a fixed threshold energy detector will lead to a large amount of detections if the threshold is too low, or to miss sound units if the threshold is too high. In this work, we propose an energy detector with an adaptive threshold that exploits some a priori knowledge of the longest/ shortest communicative sense beluga sound unit.

The algorithm is detailed as follows. If we call $x(n)$ the stochastic process at the output of the highpass filter, then a hypothesis test is that:

$$\begin{aligned} H_0: & \quad x(n) = w(n) & n = 0, 1, 2, \dots, N-1 \\ H_1: & \quad x(n) = s(n) + w(n) & n = 0, 1, 2, \dots, N-1 \end{aligned}$$

where $w(n)$ is the underlying Gaussian noise and $s(n)$ is the beluga vocalization. The value N is related to the minimum building blocks length of the beluga whale songs. If N is small enough, the stochastic processes can be considered stationary in this time interval, consequently, equations given in [17] are valid and the Neyman Pearson detector decides H_1 if

$$T(x) = \sum_{n=0}^{N-1} x(n)^2 > \gamma' \quad (1).$$

The threshold γ' can be obtained by means of Eq. (2)

$$\gamma' = \sigma^2 \cdot Q_{x_N}^{-1}(PFA_{NP}) \quad (2),$$

where $Q_{x_N}^{-1}$ is the inverse chi-square distribution with N degrees of freedom. The value σ^2 is the noise variance at the output of the filter and PFA_{NP} is the desired false alarm probability of the Neyman Pearson detector.

If the threshold γ' is maintained fixed with time, the conventional energy detector is obtained which is the optimum detector for stationary processes. However, as we have previously stated that noise process is a

quasi-stationary process. Let us call x_i the i -th N sample fragment of $x(n)$. The proposed adaptive threshold energy detector works in blocks of N samples (x_i) and the threshold γ' is recalculated from block to block according to the algorithm detailed in Fig. 2.

In the proposed algorithm, M is the length of the longest beluga vocalization with communicative sense and N is the length of the shortest beluga vocalization, σ_i^2 is the variance of the i -th fragment of acquired signal (operator $Var[\cdot]$ is used to calculate the variance) and $\Delta\alpha$ controls how fast the adaptive algorithm adapts to changes in the noise profile.

1. $\alpha = 1$, $i_0 = 1$, $i = 1$ and σ_0^2 is initialized to noise variance.
2. If the number of consecutive detections is longer than M then $\alpha = \alpha + \Delta\alpha$ and make $i = i_0$.
3. Read block x_i and calculate $\sigma_i^2 = Var[x_i]$.
4. If $\sigma_i^2 \leq (\alpha + \sigma_0^2)$ then $\sigma_0^2 = \sigma_i^2$.
5. Energy detector of block x_i according to Eq. (1) and Eq. (2) with σ_0^2 .
 - If no detection $\alpha = 1$, advance to the next block $i = i + 1$.
 - If detection:
 - If there is no detection in the previous segment, then make $i_0 = i$, save detection to classify stage and advance to the next block ($i = i + 1$).
 - If there is detection in the previous segment, then save detection to classify stage and advance to the next block ($i = i + 1$).
6. Go to step 2 (until end of $x(n)$)

FIG. 2 ADAPTIVE THRESHOLD ACTIVITY DETECTOR PROPOSED

It is important to control the magnitude of the parameter $\Delta\alpha$. If this value is very small, the algorithm follows noise variations very precisely producing a large number of false alarms when noise power changes. Thus, the convergence process becomes slow. On the other hand, if $\Delta\alpha$ is very high, the algorithm follows noise changes in larger steps and the detection probability of small amplitude vocalizations decreases. The best value for $\Delta\alpha$ has been obtained through 3000 Monte Carlo simulations of random vocalization patterns for different $\Delta\alpha$ values.

Results are shown in Fig.3 where it can be seen that $\Delta\alpha = 2.1$ maximizes the difference PD-PFA (blue curve).

Fig.4 compares (in a time varying noise scenario) a conventional fixed threshold energy detector with the

proposed adaptive threshold algorithm. In this simulation, the noise variance changes slowly compared to variance of beluga vocalizations.

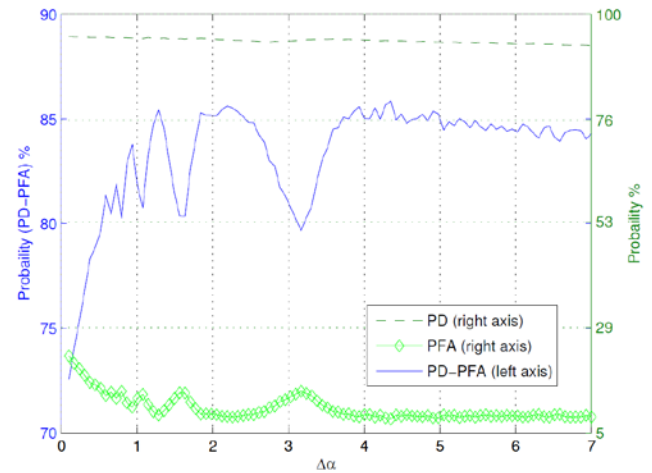


FIG. 3 DETECTION PROBABILITY (PD), FALSEALARM PROBABILITY (PFA) AND DIFFERENCE (PD-PFA) FOR THE PROPOSED ALGORITHM IN FUNCTION OF $\Delta\alpha$, ($PFA_{NP} = 10^{-4}$)

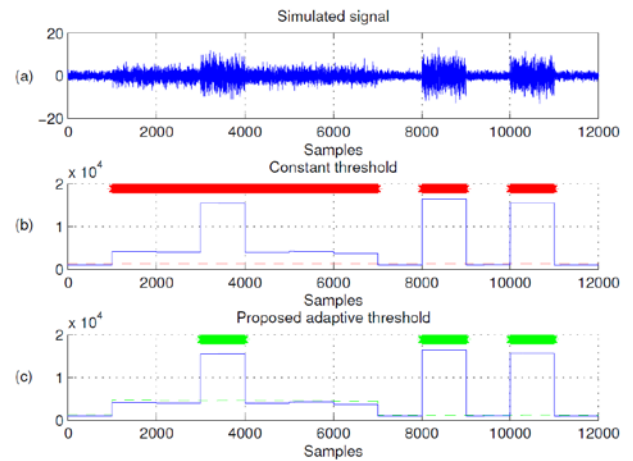


FIG. 4 COMPARISON OF THE PROPOSED ADAPTIVE THRESHOLD TO CONSTANT THRESHOLD ($M = 3 \cdot N$). BLUE CURVE: SIMULATED SIGNAL, RED DASHED CURVE: CONSTANT THRESHOLD AND GREEN DASHED CURVE: PROPOSED ADAPTIVE THRESHOLD. THE CHARACTER "***" INDICATES WHERE THERE ARE DETECTIONS

Features Description and Classifier Details

Once beluga vocalizations are detected, the obtained vocalization units must be classified. The possible classification categories related to observed animal behaviour were provided by the Oceanografic biologists based on previous works [12]. In order to enhance previous detection process, an additional category "noise" is added to the original list. The objective of this category is to remove possible noise events that resemble beluga vocalizations that might have been detected by proposed activity detector. With this category we can remove interfering sounds

such as sounds from other species or human produced noise.

TABLE 1 CLASSIFICATION CATEGORIES EMPLOYED WITH SOME SOUNDS EXAMPLES FROM THEIR REPERTOIRE AND A BRIEF DESCRIPTION RELATED TO ANIMAL BEHAVIOUR

Category	Some examples	Brief description
Tonal (Fig. 5a)	-Single tonal -Multitonal -Up-sweep whistle -Down-sweep whistle	Tonal vocalizations are typically associated to communicative behaviour
Pulsed (Fig. 5b)	-Pulsed train -Echolocation	Trains of pulses with communicative or aggressive component. As well as secholocation functions
Jawclap (Fig. 5c)	-Jaw claps	Impulsive jaw claps generally aggressive sounds
Noise	-Underwater ambient noise	Noise category to remove incorrect detection events

The categories and its main characteristics are summarized in the Table 1. Fig. 5 shows a time frequency representation sample of a sound unit belonging to each one of the categories.

A relatively small number of features have been chosen to best describe and distinguish each vocalization category. Since time-frequency representation is not needed to calculate the different features, real time processing is easily achieved with low cost processors.

The Table 2 shows the whole set of parameters and their corresponding label number. These features have been selected to adjust to different categories in order to maximize classification rate.

As it has been previously stated, some features have been selected as a simple and low computational

complexity approach to the shape of the time frequency representation. For instance, the features $[v_1, \dots, v_9]$ summarize the resonant frequencies and bandwidths of beluga multitonal vocalizations using the Fourier transform instead of the spectrogram.

TABLE 2 FEATURE SET EMPLOYED IN THE AUTOMATIC DETECTOR

Feature Number	Short Description
v_1	Fundamental Frequency f_0
v_2	Q-factor of $f_0 = \Delta f_0 / f_0$
v_3	Power spectral density of the frequency $S_x(f_0)$
v_4	First Harmonic f_1
v_5	Q-factor of $f_1 = \Delta f_1 / f_1$
v_6	Power spectral density of the frequency $S_x(f_1)$
v_7	Second Harmonic f_2
v_8	Q-factor of $f_2 = \Delta f_2 / f_2$
v_9	Power spectral density of the frequency $S_x(f_2)$
v_{10}	Skewness of the vocalization
v_{11}	Kurtosis of the vocalization
v_{12}	Autocovariance test of vocalization
v_{13}	Time reversibility measure of the vocalization
v_{14}	Voiced/ Unvoiced measure

Fig. 6 illustrates each feature definition, where f_0, f_1 and f_2 are the fundamental frequency, first and second harmonic, respectively. The -3 dB bandwidths $\Delta f_0, \Delta f_1$ and Δf_2 are employed to obtain the normalized bandwidth or Q-factor (features v_2, v_5 and v_8) according to equations given in Table 2. Power spectral amplitude of the three main frequencies is also obtained (features v_3, v_6 and v_9).

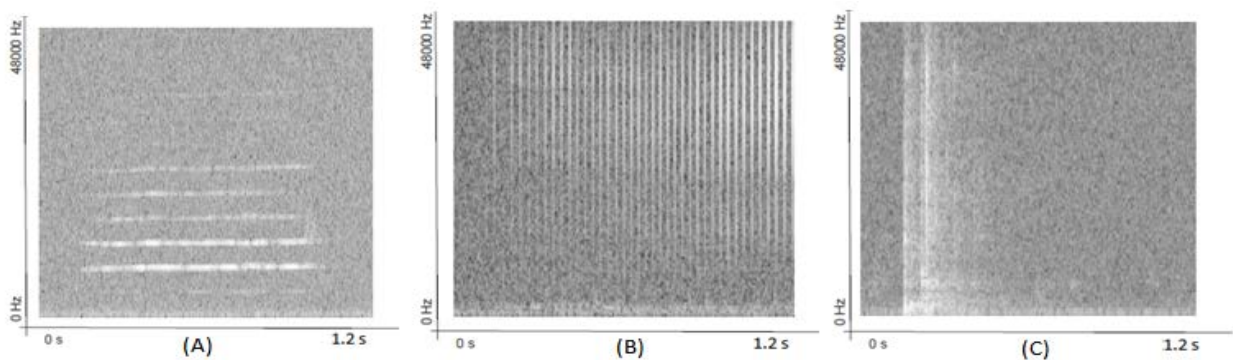


FIG. 5 TIME FREQUENCY REPRESENTATIONS OF: A) PURE TONAL VOCALIZATION (MULTITONAL), B) PURE PULSED VOCALIZATION AND C) JAWCLAP VOCALIZATION

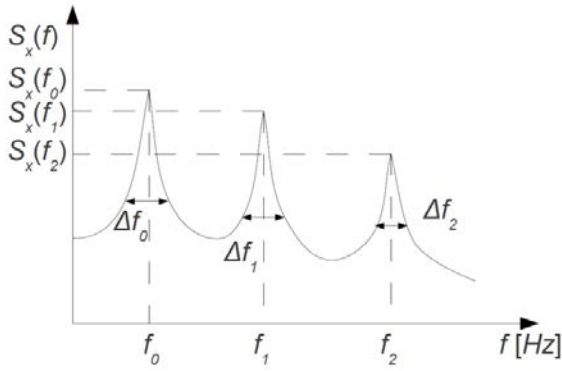


FIG. 6 DEFINITION OF FREQUENCY RELATED FEATURES (v_1 TO v_9)

Other parameters are related to higher order statistics of the vocalization (features v_{10} , v_{11} , v_{12} and v_{13}). The preliminary study carried out by the researchers showed that this statistical information could be useful to identify some particular sound units. The feature v_{10} computes the skewness of the sound unit. If the vocalization is regarded as an stochastic process, the skewness, a measure of the asymmetry of the probability distribution, is computed as described in Eq. (3).

$$v_{10} = E \left[\left(\frac{x - \mu}{\sigma} \right)^3 \right] \approx \frac{1/N \sum_{n=1}^N (x(n) - \bar{x})^3}{(1/N \sum_{n=1}^N (x(n) - \bar{x})^2)^{3/2}} \quad (3)$$

The operator $E[\cdot]$ is the expected value operator and \bar{x} is the arithmetic average. The kurtosis (feature v_{11}), a measure of the "peakedness" of the probability distribution, is obtained by sample averaging as described in Eq. (4).

$$v_{11} = E \left[\left(\frac{x - \mu}{\sigma} \right)^4 \right] - 3 \approx \frac{1/N \sum_{n=1}^N (x(n) - \bar{x})^4}{(1/N \sum_{n=1}^N (x(n) - \bar{x})^2)^2} - 3 \quad (4)$$

A couple of parameters that give information about the vocalization nonlinearity have also been added in the feature set: third-order autocovariance and time reversibility test of the vocalization (parameters v_{12} and v_{13} of the feature set).

The so called "third-order autocovariance" [18, 20] is a higher-order extension of the traditional autocovariance. This parameter was computed as described in [2] (see Eq. (5)).

$$v_{12} = t^c = E[x(n)x(n-1)x(n-2)] \quad (5)$$

The time reversibility parameter (v_{13}) is a sample estimate of the slope skewness normalized by the estimate of slope standard deviation to the third power ($\hat{\sigma}^3$) [22]. A statistical process is said to possess time-reversibility if statistical properties are identical

when both forwards and backwards in time are examined. This feature is computed as:

$$v_{13} = \frac{1}{\hat{\sigma}^3} \sum_{n=1}^N \left(\frac{x(n) - x(n-1)}{T_s} \right)^3 \quad (6)$$

For times series that exhibit time-reversibility, it is expected $v_{13} \approx 0$. In contrast, processes that are time irreversible yield values of $v_{13} > 0$ or $v_{13} < 0$. Both parameters, v_{12} and v_{13} , are typically employed to detect nonlinearities in time series [14].

Finally, an additional parameter v_{14} was added to the set. This parameter is inspired by linear prediction coding algorithms typically employed in the analysis and synthesis of human speech to identify voiced or unvoiced sound units [16, 10]. In order to compute this parameter, the auto-correlation of the linear prediction error (R_{ee}) and the pitch period ($T_p = .2$ s) of the beluga vocalization have been estimated. The auto-correlation is evaluated at the estimated pitch period as seen in Eq. (7). This value is normalized by the vocalization energy and the factor $(1 - T_p/N)$ to compensate the triangular decay of the autocorrelation estimate.

$$v_{14} = \frac{1}{1 - T_p/N} \cdot \frac{R_{ee}(T_p)}{R_{ee}(0)} \quad (7)$$

Classification in a Controlled Environment

The proposed algorithm has been tested with real beluga signals recorded in the Oceanographic of Valencia. Details are given as follows.

Beluga whale vocalizations were recorded with a sound acquisition system Roland (Edirol) FA-101 (24 bits and frequency sample $f_s = 96$ KHz), a Bruel&Kjaer 8103 hydrophone (0.1 Hz-180 KHz) and a Bruel&Kjaer 2692 Nexus amplifier (0.1 Hz -100 KHz). These audio files were joined into a single one containing all the different vocalizations. The Oceanographic biologist listened and classified each one of the acoustic events (echolocation pulses were excluded). The aforementioned file contains 560 vocalizations derived from two different specimens: the female beluga (Yulka) and the male (Cairo).

The described activity detector has been applied to this test file with the following settings. The longest communicative sense beluga vocalization is chosen to be $M = 1.5$ sec. and the shortest is $N = .1$ sec. The detection algorithm was set to work with $\Delta_\alpha = 2.1$ and $PFA_{NP} = 10^{-4}$. With these settings, the detection percentages and false alarm rate were

obtained by means of comparison of the output of the proposed detector with the biologist detections. The achieved percentages at the output of the detector are $PD = 98.1\%$ and $PFA = 37.6\%$. The slightly higher value of the PFA is due to the fact that the algorithm settings are fixed, so that no beluga sound unit is missed (according to the biologist manual classification). The sound units detected were fed into the proposed classifier.

A simple Naive Bayes classification algorithm was used to obtain the recognition rate [6]. The training set and test set used were proportionate from data base in [12]. Both of them have the same vocalization proportions roughly.

The overall correct classification percentage was 88.3%. However, this percentage was not equally distributed among all categories presented. Table 3 shows the recognition rate per category and its confusion matrix. As it can be appreciated, the average classification percentage for the noise category allows rejecting almost every undesired detection produced at the detection stage. This fact can be exploited by the proposed system to achieve very low beluga sound losses, lowering the detection thresholds at the detection stage.

TABLE 3 CONFUSION MATRIX. AVERAGE RECOGNITION RATE PER CATEGORY (GREY)

Category	Tonal	Pulsed	Jawclap	Noise
Tonal	83.7%	14.8%	1.1%	0.4%
Pulsed	16.1%	79.9%	2.4%	1.6%
Jawclap	0%	6.4%	88.2%	6.4%
Noise	0%	0%	3.9%	96.1%

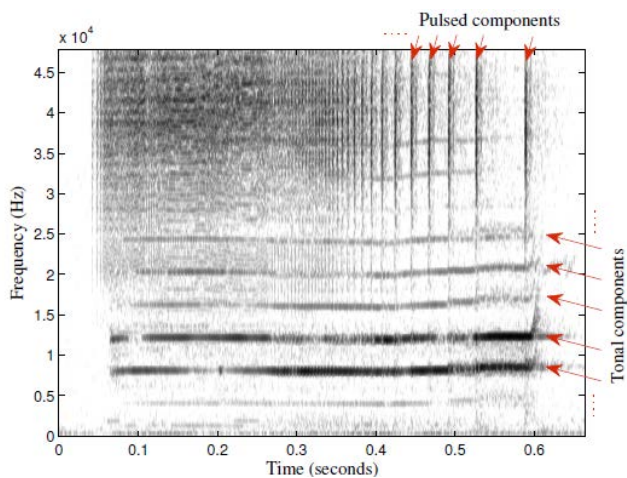


FIG. 7 MIXED (PULSED AND TONAL COMPONENTS) BELUGA SOUND

The best achieved classification rate was obtained for the jawclap category. However, so high classification

percentages were not achieved by tonal and pulse categories. A possible explanation is due to the fact that beluga whales can produce sounds with both components (pulsed and tonal) mixed, which can be seen in Fig. 7. The proposed algorithm may fail and classify this sound unit as the predominant component (tonal or pulsed), which may not to match the biologist criteria.

Nonlinearity Measure of Some Beluga Sounds

It has been showed in [23] that nonlinear production mechanisms allow individuals to generate highly complex and unpredictable vocalizations without requiring equivalently complex neural control mechanisms. In [4], the presence of nonlinearities was observed and measured for humpback whales. In [13], qualitative descriptions and quantitative analyses of nonlinearities in the vocalizations of killer whales (*Orcinus orca*) and North Atlantic right whales (*Eubalaena glacialis*) were provided.

Some of the features proposed in Table 2 can be used in a similar way to automatically detect the presence of nonlinearities in certain vocalizations of the beluga whales. In order to show this, we have represented in Fig. 8 the values of the feature $|v_{13}|$ (time reversibility measure) compared with the feature v_{14} (voiced/unvoiced parameter) obtained with a sound register containing 313 vocalizations of all the categories described in Table 1. As it can be seen, most jawclap as well as some tonal and pulsed vocalizations seem to be produced by nonlinear mechanisms (non-zero time reversibility values). However, a more detailed study should be done to assert that high values in the nonlinear indicators really come from nonlinearities.

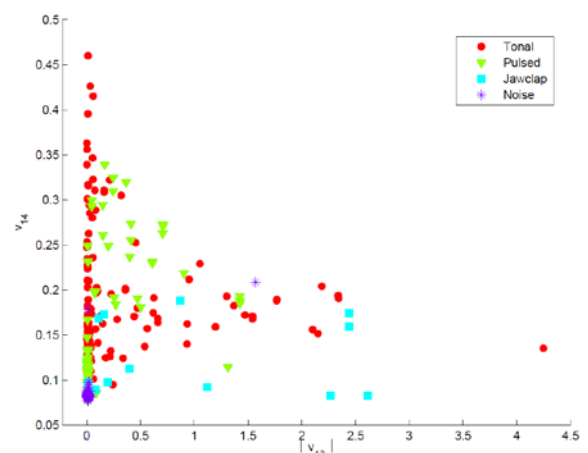


FIG. 8 FEATURE $|v_{13}|$ (TIME REVERSIBILITY MEASURE) OF DIFFERENT CATEGORIES OVER FEATURE v_{14} (VOICED/UNVOICED PARAMETER)

A surrogate test has been done to measure the nonlinearity degree. This is a procedure in which a given metric is evaluated in the original data and in a surrogate version of the data artificially generated to resemble the original data.

The method employed for surrogate generation, the Iterative Amplitude Adapted Fourier Transform (iAAFT) [21], can be employed for stationary time series even when the data does not follow Gaussian distribution. The surrogate data generated have both amplitude spectrum and signal statistical distribution matched to the original data.

As a consequence of non-stationarity of beluga sound units, the proposed surrogate algorithm has to be employed only with those sound fragments where stationarity holds. Thus, only small fragments of 10ms of the tonal category of beluga whale sounds are going to be processed using this technique. Otherwise, surrogate data generation does not preserve non-stationarity and differences in the score values will be due to non-stationarity rather than nonlinearity. Additionally, using only tonal vocalizations, an indication can be acquired for beluga whales of irregular periodic vibrations similar to those produced in terrestrial mammals [7].

The tonal vocalizations employed to perform the nonlinearity measure are named: Whistle Creak, Creak Whistle and Flat Whistle. In Fig. 9, the spectrogram of these sound units can be seen. Inspection of the sounds reported evidences of nonlinear behaviour. This was done by inspecting their time frequency representation to look for frequency jumps and subharmonics. The selected features (metrics) for the surrogate test are v_{12} and v_{13} (both features are typically employed to detect nonlinearities in time series) [16]. The procedure description is as follows: firstly a significant set of surrogate series is artificially generated with the aforementioned algorithm, then statistics sensitive to nonlinearity (v_{12} and v_{13}) are determined on both the surrogate and the original time series, and finally the values are compared by means of a rank test. The position index (rank) of each feature with respect to the surrogates is determined.

The percentage rank ($Pr[\%]$) has been calculated for several regions inside a given vocalization (red markers in Fig. 9). The results, obtained for 200 surrogates in each test, can be seen in Table 4.

The results of Table 4 show that if a short time window analysis of tonal vocalizations is employed

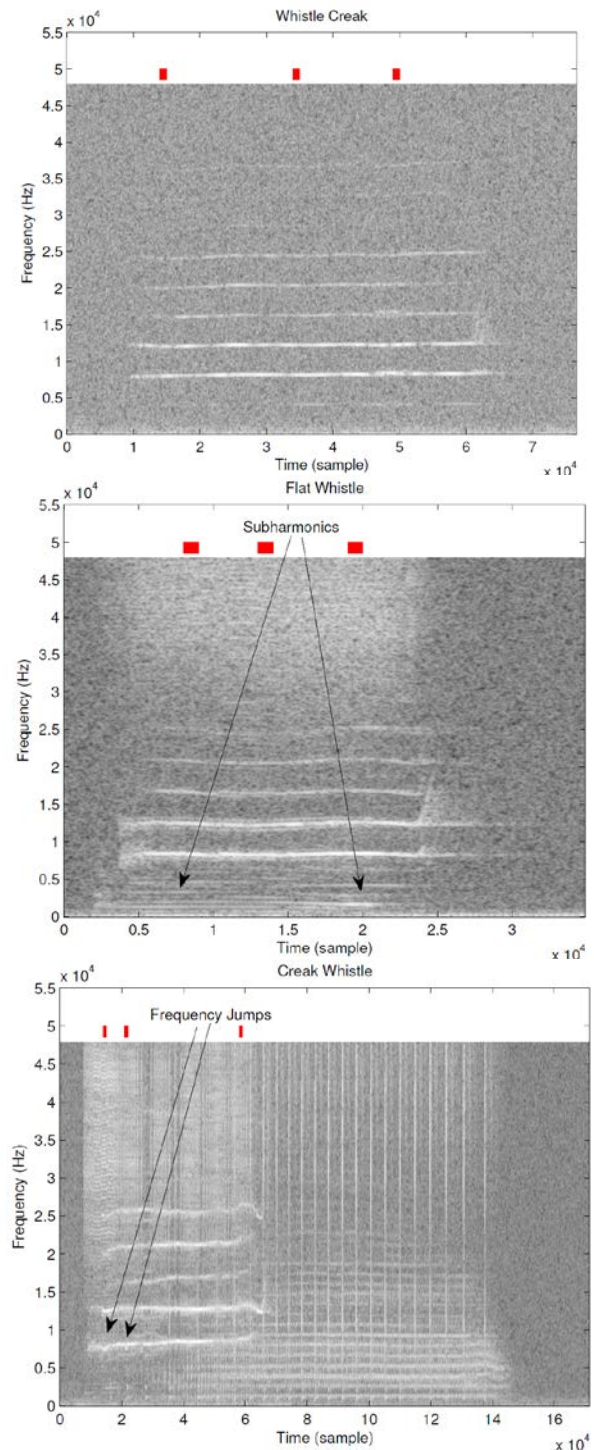


FIG. 9 TIME-FREQUENCY REPRESENTATIONS OF DIFFERENT TONAL BELUGA SOUND UNITS. THE RED MARKERS INDICATE THE REGIONS WHERE THE NONLINEARITY RANK INDEX FOR THE PRESENTED METRICS HAS BEEN COMPUTED. THE ARROWS ARE DIRECTED TO OBSERVED NONLINEARITIES

the high values of v_{12} and v_{13} indicate the presence of nonlinearities in the beluga whale vocalizations. These results can be used to devise automatic tests to look for subharmonics and frequency jumps in beluga sounds. The study of nonlinearities in beluga

vocalizations could be a key factor in the study of the beluga anatomy. For example, the presence of nonlinearities in the vocalizations could indicate that some of the complex tonal beluga whales sounds may be generated by irregular vibration of some sort of fold similar to vocal folds in terrestrial mammals [7]. However, it is important to highlight that not all vocalizations that appear in Fig. 9 and exhibit a high v_{13} value are due to nonlinear dynamics. This feature calculated at the classification analysis for the whole vocalization length may indicate non-stationarity as well. It will be important then when developing tests for nonlinear detection to guarantee the stationarity of the whale song fragment or to use an alternative surrogate data generation that could be employed in non-stationarity signals [1, 19].

TABLE 4 PERCENTAGE RANK (Pr[%]) FOR DIFFERENT SOUNDS. THE NUMBER OF SURROGATE DATA EMPLOYED IS 200. THE TABLE REFERS TO THE ANALYSIS PERFORMED AT THE POSITIONS OF THE 1ST, 2ND AND 3RD RED MARKERS OF FIG. 8

Sound name	Rank test	1st	2nd	3rd
Whistle Creak	v_{12}	30%	71%	73%
	v_{13}	63%	36%	56%
Creak Whistle	v_{12}	99%	100%	44%
	v_{13}	89%	99%	50%
Flat Whistle	v_{12}	96%	29%	94%
	v_{13}	49%	71%	68%

Conclusions and Future Work

An automatic system has been developed for beluga sound detection and classification. The proposed algorithm has been designed to work in real time with moderate computational requirements.

The detection algorithm is based on an adaptive threshold energy detector whereas the classification is performed with a Naive Bayes classifier. This work demonstrates that feature extraction for classification and machine learning is a feasible alternative to be applied instead of the typical approaches based on pattern recognition of time frequency representations of cetaceous sounds. The proposed detection algorithm although its simplicity, gives good results in simulations and real signals (98% of detection percentage).

The results, according to the classification proposed by biologists, are divided into three categories: tonal sounds (communicative), pulsed sounds (communicative and aggressive) and jawclaps (aggressive). Despite the complex nature of some beluga sounds (with mixed tonal and pulsed components), the classification percentages achieved

are quite high. The averaged classification percentage of the system is close to 88%.

On the other hand, this work illustrates that feature extraction can be a useful tool to design automatic detectors of irregularities in beluga whale vocalizations. Nonlinearities in beluga whale songs can be quantitatively characterized by means of comparing nonlinear metrics with surrogate data. This may lead to indicators that are capable to automatically detect and characterize “nonlinear dynamics” in beluga bioacoustics. Frequency jumps and subharmonics detections are given as example.

Although promising results have been obtained, the proposed approach has to be tested in open sea and for different cetacean species. The authors have started to work on these topics as well as many other related to characterization and signal modality of cetacean sounds.

ACKNOWLEDGMENT

This work has been supported by the national R + D program under Grant TEC2011-23403 (Spain), and appreciation should be paid to the Càtedra Telefónica in the Universitat Politècnica de València.

REFERENCES

- A. Schmitz, and T. Schreiber, “Surrogate data for Non-Stationary signals”, in Workshop on Chaos in Brain?, edited by P. G. K. Lehnertz, J. Arnhold and E. S. W. S.C. E. Elger, 222-225 (1999).
- C. Nikias, and A. Petropulu, Higher-Order Spectra Analysis a Nonlinear Signal Processing Framework, pp. 123-138 (Prentice Hall, Englewood Cliffs, New Jersey) (1993).
- D. K. Mellinger, and C. W. Clark, “Recognizing transient low frequency whale sounds by spectrogram correlation”, J. Acoust. Soc. Am. 107, 3518-3529 (2000).
- E. Mercado, J. Schneider, A. Pack, and L. Herman, “Sound production by singing humpback whales”, J. Acoust. Soc. Am. 127, 2678-2691 (2010).
- H. Figueroa, “Xbat v.5”, Cornell University Bioacoustics Research Program <http://xbat.org/> (date last viewed 12/5/2011).
- I. Rish, “An empirical study of the naive Bayes classifier”, IJCAI Workshop on Empirical Methods in Artificial Intelligence (2001).
- I. Tokuda, T. Riede, J. Neubauer, M. Owren, and H. Herzel,

- "Nonlinear analysis of irregular animal vocalizations", J. Acoust. Soc. Am. 111, 2908-2919 (2002).
- J. Ramirez, J. M. Gorriz, and J. C. Segura, "Voice Activity Detection. Fundamentals and Speech Recognition System Robustness, Robust Speech Recognition and Understanding", Michael Grimm and KristianKroschel (Ed.), ISBN: 978-3-902613-08-0, InTech. (2007).
- K. Stafford, C. Fox, and D. Clark, "Long-range acoustic detection and localization of blue whale calls in the northeast pacific ocean", J. Acoust. Soc. Am. 3616-3625(1998).
- L. Rabiner, M. Cheng, A. Rosenberg, and C. McGonegal, "A comparative performance study of several pitch detection algorithms", Acoustics, Speech and Signal Processing, IEEE Transactions on 24, 399-418 (1976).
- M. Azorin, M. Castellote, and J. Esteban, "Birth prediction using acoustics in captive beluga whales", in 19th International Congress on Acoustics (Madrid, Spain), 1-6 (2007).
- M. Castellote, and F. Fossa, "Measuring acoustic activity as a method to evaluate welfare in captive beluga whales (*delphinapterusleucas*)", Aquatic Mammals 32, 325-333(2006).
- R. B. Tyson, D. P. Nowacek, and P. J. O. Miller, "Nonlinear phenomena in the vocalizations of North Atlantic right whales (*Eubalaenaglacialis*) and killer whales (*Orcinus orca*)", J. Acoust. Soc. Am., Volume 122, Issue 3, pp. 1365-1373 (2007).
- R. Miralles, L. Vergara, A. Salazar, and J. Igual, "Blind detection of nonlinearities in multiple echo ultrasonic signals", IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control 55,3, 1-11 (2008).
- R. S. Payne, and S. McVay, "Songs of humpback whales", Science 173, 585-597 (1971).
- R. W. Schafer, and L. R. Rabiner, "System for automatic formant analysis of voiced speech", J. Acoust. Soc. Am. 47, 634-648 (1970).
- S. M. Kay, Fundamentals of Statistical Signal Processing: Detection Theory (Volume II), pp. 142-146, 1 edition (Prentice Hall PTR, Upper Saddle River, New Jersey) (1993).
- T. Gautama, M. V. Hulle, and D. Mandic, "On the characterisation of the deterministic/ stochastic and linear/ nonlinear nature of time series", Technical Report DPM-04-5, Imperial College London (2004).
- T. Nakamura, and M. Small, "Small-shuffle surrogate data: Testing for dynamics in fluctuating data with trends", Phys. Rev. E 72, 056 216-1-056 216-6 (2005).
- T. Schreiber and A. Schmitz, "Discrimination power of measures for nonlinearity in a time series", Phys. Rev. E55, 5443-5447 (1997).
- T. Schreiber and A. Schmitz, "Improved surrogate data for nonlinearity tests", Physical Review Letters 77, 635.638 (1996).
- T. Schreiber, and A. Schmitz, "Testing for nonlinearity inunevenly sampled time series", Phys. Rev. E 59, 4044-4048 (1999).
- W. Fitch, J. Neubauer, and H. Herzel, "Calls out of chaos: the adaptive significance of nonlinear phenomena in mammalian vocal production", Animal behaviour 63,407-418 (2002).



R. Miralles was born in Valencia (Spain) in 1971. He received the Bachelor of Science in Engineering in 1995 and the Ph.D. degree in Telecommunication Engineering in 2000 from the Universidad Politécnica de Valencia (UPV). In 1996 he became a lecturer in the Escuela Politécnica Superior de Gandia (Valencia). Since 2000, he has been working as an Assistant Professor in the Escuela Tecnica Superior de Ingenieros de Telecomunicación (Valencia). He is responsible for the development of algorithms and systems for nondestructive testing in the alimentary industry using ultrasound and for passive acoustic detection of cetaceans. He is member of the Institute of Telecommunication and Multimedia Applications (iTEAM) of UPV (Board of Directors).

His research interests are higher order statistics, nonlinear signal processing, signal processing applications for ultrasonic systems and bioacoustic signal processing. He has published 55 papers in journals and conference communications.



G. Lara was born in Valencia. He has recieved the Telecommunication Engineering degree from the Universidad Politécnica de Valencia (UPV) Spain in 2010. He is a Ph.D. student in the Institute of Telecommunication and Multimedia Applications (iTEAM) of UPV.

His research interest is focused on pattern recognition and statistical processing applied at the submarine acoustic.

Currently, he involves in the development of a submarine buoy capable of recording cetaceans sounds without losing samples, programming the electronics and the internal hardware.



A. Carrión was born in Lorca (Spain). She received the Telecommunication Engineering degree from the Universidad Politécnica de Valencia (UPV) in 2011, and her Master Thesis was carried out at Fraunhofer Institute IOSB, Karlsruhe (Germany). Currently, she is a Ph.D. student in the Institute of Telecommunication and Multimedia Applications (iTEAM) of UPV. Her research interests include nonlinear signal

processing and signal modality characterization.

J.A. Esteban was born in Valencia. He is Coordinator of the Research Department of the Oceanogràfic of the City of Arts and Science in Valencia. He holds master degree in Environmental Management and a bachelor in Biology, both from Valencia University in Valencia, (Spain). His research is focused on cetacean bioacoustics, with the animal collection of the Oceanogràfic marine mammals and with Mediterranean cetacean wildlife as well. He is currently working on long term monitoring studies of fin whales seasonal patterns in the Spanish Mediterranean coast using passive acoustic monitoring systems and seasonal patterns of local bottlenose dolphins populations using T-POD devices.